



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Cognitive development Respiratory Tract Illness and Effects of eXposure (CORTEX) project: Combining high spatial resolution pollution measurements with individual level data, a methodological approach.

Citation for published version:

Rodgers, S, Lyons, J, Mizen, A, Berridge, D, Akbari, A, Carruthers, D, Davies, G, Dearden, L, Doherty, R, Lake, I, Mavrogianni, A, Milojevic, A, Strickland, S & Wilkinson, P 2018, 'Cognitive development Respiratory Tract Illness and Effects of eXposure (CORTEX) project: Combining high spatial resolution pollution measurements with individual level data, a methodological approach.', *International Journal of Population Data Science*, vol. 3, no. 4. <https://doi.org/10.23889/ijpds.v3i4.802>

Digital Object Identifier (DOI):

[10.23889/ijpds.v3i4.802](https://doi.org/10.23889/ijpds.v3i4.802)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

International Journal of Population Data Science

Publisher Rights Statement:

September 2018 c The Authors. Open Access under CC BY-NC-ND 4.0

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



International Journal of Population Data Science

Journal Website: www.ijpds.org



Cognitive development Respiratory Tract Illness and Effects of eXposure (COR-TEX) project: Combining high spatial resolution pollution measurements with individual level data, a methodological approach.

Rodgers, S¹, Lyons, J², Mizen, A², Berridge, D², Akbari, A³, Carruthers, D⁴, Davies, G¹, Dearden, Institute of Fiscal Studies⁵, Doherty, R⁶, Lake, I⁷, Mavrogianni, A⁸, Milojevic, A⁹, Strickland, S⁴, and Wilkinson, P⁹

¹Swansea University Medical School

²Farr Institute, Swansea University Medical School

³Health Data Research UK - Wales and Northern Ireland, Swansea University Medical School

⁴Cambridge Environmental Research Consultants

⁵Institute of Fiscal Studies

⁶University of Edinburgh

⁷University of East Anglia

⁸University College London

⁹London School of Hygiene and Tropical Medicine

Introduction

The Secure Anonymised Information Linkage (SAIL) databank facilitated linkage of routinely collected health and education data, high spatial resolution pollution modelling and daily pollen measurements for 18,241 pupils in 7 cross-sectional cohorts across Cardiff city, UK, to investigate effects of air quality and respiratory health conditions on education attainment.

Objectives and Approach

An urban atmospheric dispersion and chemistry modelling system (ADMS-Urban) simulated modelled hourly concentrations of air pollutants: PM_{2.5}, PM₁₀, NO₂ and ozone levels. These were summarised into minimum, average and maximum daily readings for 4 time periods (e.g. school hours 9am-3pm) for all home and school locations across Cardiff between 2009 and 2015. The combination of different pollutants, measurements and time-periods created a comprehensive multi-row dataset per location. We transformed the dimensionality of this high-resolution data to create one row of summarised data per pupil per cohort, in preparation for statistical analysis.

Results

157,361 school and home locations across Cardiff were anonymised and household linkage fields were appended to combine pollution estimates at the household/school to individual health data. The pollution dataset contained 369 columns, 472,083 rows per year with one column per location, pollutant type, pollutant measurement, daily time-period, and

day of year. Dataset transformation reduced algorithm computation by creating a single date column, producing a five column, 3,446,205,900-row matrix per year dataset. The algorithm adjusted for weekends, school/bank holidays and allowed location to vary 3pm-5pm on school days when pupil location was uncertain. The algorithm calculated tailored pollution exposures per pupil for revision and examination periods, creating one row per pupil and reducing 7 years of data and 24 billion rows to 18,241.

Conclusion/Implications

We successfully linked 95% of the cohorts' household/school pollution data to their corresponding health and education data. This demonstrates data linking retrospective exposures for total populations using multiple daily locations, and extends our analysis platform for natural experiments to include daily exposure. Future work includes adding modelled route exposures.

